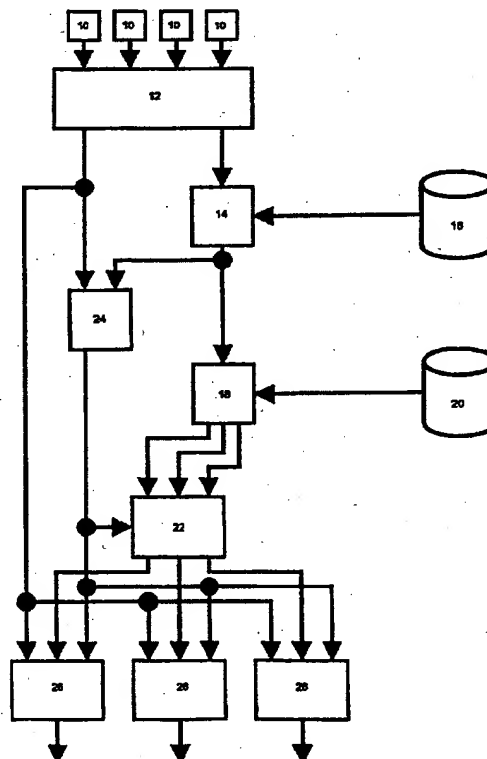


<p>(51) International Patent Classification <sup>6</sup> :  <b>H04N 5/272, 5/262</b></p>	<p><b>A1</b></p>	<p>(11) International Publication Number: <b>WO 97/03517</b>          (43) International Publication Date: <b>30 January 1997 (30.01.97)</b></p>
<p>(21) International Application Number: <b>PCT/GB96/01682</b>          (22) International Filing Date: <b>15 July 1996 (15.07.96)</b>          (30) Priority Data: <b>9514313.7 13 July 1995 (13.07.95) GB</b>          (71)(72) Applicant and Inventor: <b>BEATTIE, Robert, John [GB/GB]; 47 Stewartfield Drive, Eastkilbride G74 4UA (GB).</b>          (74) Agent: <b>MURGITROYD &amp; COMPANY; 373 Scotland Street, Glasgow G5 8QA (GB).</b></p>		<p>(81) Designated States: <b>AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</b></p> <p><b>Published</b>  <i>With international search report.          Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>

(54) Title: METHODS AND APPARATUS FOR PRODUCING COMPOSITE VIDEO IMAGES

**(57) Abstract**

A system for automatically generating and adding secondary video images (such as advertising material) to primary video images or real world scenes (such as a live sports event) in such a way that the secondary image appears to be physically present in the scene represented by the primary image when the composite image is viewed subsequently. A "live" image from one of a number of cameras (10), is selected by an editing desk (12) for transmission. Prior to transmission, a secondary image is selected from a database (20) for inclusion in the final image, such that it appears superimposed on a physical target space in the first image. The selected image is transformed in terms of size, shape, orientation and lighting effects before being combined with the primary image. The transformation is based on a computed "expected image", which is derived from a computer model (16) of the environment containing the first image (such as a sports arena) and data transmitted from the camera regarding its location, orientation, focal length, etc. The expected image is matched with the first image in a matching module (24) to refine the alignment of the computed target space with the actual target space, and to identify lighting variations and foreground objects in the first image and apply these to the second image as seen in the final composite image. Multiple composite images may be generated including different secondary images so that, for example, different advertisements can be included in different composite images for transmission to different audiences.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LJ	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

1     **"Methods and Apparatus for Producing Composite Video**  
2     **Images"**

3  
4     The present invention relates to a system for  
5     automatically generating and adding secondary images to  
6     primary images of real world scenes in such a way that  
7     the secondary image appears to be physically present in  
8     the scene represented by the primary image when the  
9     composite image is viewed subsequently.

10  
11    It is particularly envisaged that the invention be  
12    applied to the presentation of advertising material  
13    (secondary images) within primary images including, but  
14    not limited to, television broadcasts, video  
15    recordings, cable television programmes and films. It  
16    is applicable to all video/TV formats, including  
17    analogue and digital video, PAL, NTSC, SECAM and HDTV.  
18    This type of advertising is particularly applicable to,  
19    but is not limited to, live broadcasts of sports  
20    events, programmes of highlights of sports events,  
21    videos of sports events, live broadcasts of important  
22    state events, television broadcasts of "pop" concerts  
23    etc.

24  
25    Prior practice relating to the placement of

1 advertisements within scenes represented in TV/video  
2 images includes:

3 physical advertising hoardings which can be placed  
4 at appropriate places in a scene or venue such that  
5 they sometimes appear in the images; such hoardings can  
6 be either simple printed signs or electromechanical  
7 devices allowing the display of several fixed  
8 advertisements consecutively;

9 advertisements which are placed directly onto  
10 surfaces within the scene, for example, by being  
11 painted onto the outfield at a cricket match, or by  
12 being placed on players' clothes or by being painted  
13 onto racing car bodies;

14 small fixed advertisements, for example, company  
15 logos, which are simply superimposed on the image of  
16 the scene.

17

18 These methods have the following disadvantages:

19 each physical advertising hoarding can present, at  
20 most, a few static images; it cannot be substantially  
21 varied during the event, nor can its image be changed  
22 after the event other than by a painstaking manual  
23 process of editing individual images;

24 advertisements made, for example, on playing  
25 surfaces or on participants clothing, have to be  
26 relatively discreet otherwise they intrude too much  
27 into the event itself;

28 fixed advertisements, such as company logos,  
29 superimposed on the image, look artificial and  
30 intrusive since they are obviously not part of the  
31 scene being viewed.

32

33 The present invention concerns a system whereby  
34 secondary images, such as advertising material, can be  
35 combined electronically with, for example, a live  
36 action video sequence in such a manner that the

1 secondary image appears in the final composite image as  
2 a natural part of the original scene. For example, the  
3 secondary image may appear to be located on a hoarding,  
4 while the hoarding in the original scene contains  
5 different material or is blank. This allows, for  
6 example, different advertising material to be  
7 incorporated into the scene to suit different broadcast  
8 audiences.

9  
10 Numerous systems exist for combining video images for  
11 various purposes. The prior art in this field includes  
12 the use of "colour keying" (also known as "chroma  
13 keying") in which a foreground object, such as a  
14 weather forecaster, is in front of a uniform background  
15 of a single "key" colour. A second video source  
16 provides another signal, such as a weather map. The  
17 two video signals are mixed together so that the second  
18 video signal replaces all parts of the first video  
19 signal which have the key colour. A similar approach  
20 is employed in "pattern-keying". Alternatively, of  
21 course, individual frames of the primary image could be  
22 edited manually to include the secondary image.

23  
24 It has previously been proposed to use video systems of  
25 this general type to insert advertising material into  
26 video images, one example being disclosed in  
27 WO93/02524. WO93/06691 discloses a system having  
28 similar capabilities.

29  
30 Colour keying works well in very restricted  
31 circumstances where the constituent images can be  
32 closely controlled, such as in weather forecasting or  
33 pre-recorded studio productions. However, it does not  
34 work in the general case where it is desired to mix  
35 unrestricted background images in parts of unrestricted  
36 primary images. The same applies generally to pattern-

1     keying systems. Replacing physical advertising signs  
2     by manually editing series of images is not feasible  
3     for live broadcasts and is extremely costly even for  
4     use with recorded programmes.

5  
6     Existing systems such as these are not well suited for  
7     the purposes of the present invention. Even where  
8     prior proposals relate specifically to the insertion of  
9     advertising material in video images, such proposals  
10    have not addressed one or more issues such as coping  
11    with foreground objects or with lighting effects or  
12    with multiple cameras.

13  
14    In accordance with a first aspect of the present  
15    invention there is provided a method of modifying a  
16    first video image of a real world scene to include a  
17    second video image, such that said second image appears  
18    to be superimposed on the surface of an object  
19    appearing within said first image, wherein said second  
20    image is derived by transforming a preliminary second  
21    image to match the size, shape and orientation of said  
22    surface as seen in said first image and said second  
23    image is combined with said first image to produce a  
24    composite final image;

25       said method including:

26       a preliminary step of constructing a three-  
27    dimensional computer model of the environment  
28    containing the real world scene, said model including  
29    at least one target space within said environment upon  
30    which said second image is to be superimposed;

31       generating camera data defining at least the  
32    location, orientation and focal length of a camera  
33    generating said first image; and

34       transforming the preliminary second image on the  
35    basis of said model and said camera data so as to match  
36    said target space as seen in the first image, prior to

1 combining said first image and said second image.

2

3 In accordance with a second aspect of the invention  
4 there is provided apparatus for generating a composite  
5 video image comprising a combination of a first video  
6 image of a real world scene and a second video image,  
7 such that said second image appears to be superimposed  
8 on the surface of an object appearing within said first  
9 image, including:

10 at least one camera for generating said first  
11 image;

12 means for generating said second image by  
13 transforming a preliminary second image to match the  
14 size, shape and orientation of said surface as seen in  
15 said first image; and

16 means for combining said second image with said  
17 first image to produce a composite final image;

18 said apparatus including:

19 means for storing a three-dimensional computer  
20 model of the environment containing the real world  
21 scene, said model including at least one target space  
22 within said environment upon which said second image is  
23 to be superimposed;

24 means for generating camera data defining at least  
25 the location, orientation and focal length of a camera  
26 generating said first image; and

27 means for transforming the preliminary second  
28 image on the basis of said model and said camera data  
29 so as to match said target space as seen in the first  
30 image, prior to combining said first image and said  
31 second image.

32

33 Further aspects and preferred features of the invention  
34 are defined in the Claims appended hereto.

35

36 Embodiments of the invention will now be described, by

1 way of example only, with reference to the accompanying  
2 drawing, which is a schematic block diagram of a system  
3 embodying the invention.

4  
5 The overall scheme of the invention is illustrated in  
6 the drawing. One or more cameras 10 are deployed to  
7 provide video coverage of an event in a venue, such as  
8 a sporting arena (not shown). The following discussion  
9 relates particularly to "live" coverage, but it will be  
10 understood that the invention is equally applicable to  
11 processing pre-recorded video images and associated  
12 data.

13  
14 Each of the cameras 10 is augmented by the addition of  
15 a hardware module (not shown) adapted to generate  
16 signals containing additional data about the camera,  
17 including position and viewing direction in three  
18 dimensions, and lens focal length. A wide variety of  
19 known devices may be used for providing data about the  
20 orientation of a camera (e.g. inclinometers,  
21 accelerometers, rotary encoders etc.), as will be  
22 readily apparent to those of ordinary skill in the art.

23  
24 The video signal from each camera 10 in operation at a  
25 particular event is passed to an editing desk 12 as  
26 normal, where the signal to be transmitted is selected  
27 from among the signals from the various cameras.

28  
29 The additional camera data is passed to a modelling  
30 module (computer) 14 which has access to a predefined,  
31 digital 3-d model of the venue 16. The venue model 16  
32 contains representations of all aspects of the venue  
33 which are significant for operation of the system,  
34 typically including the camera positions and the  
35 locations, shapes and sizes of prominent venue features  
36 and all "target spaces" onto which secondary images are



1 to be superimposed by the system, such as physical  
2 advertising hoardings.

3  
4 The modelling module 14 uses the camera location,  
5 orientation and focal length data to compute an  
6 approximation of the image expected from the camera 10  
7 based on transformed versions of items forming part of  
8 the model 16 which are visible in the camera's current  
9 view.

10  
11 The modelling module 14 also calculates a pose vector  
12 relative to the camera view vector for each of the  
13 target spaces visible in the image. Target spaces into  
14 which the system is required to insert secondary images  
15 are referred to herein as "designated targets".

16  
17 The additional camera data is also passed to the  
18 secondary image generation module 18 which generates a  
19 preliminary secondary image for each designated target  
20 in the primary image. A library of secondary images is  
21 suitably stored in a secondary image database 20,  
22 accessible by the secondary image generation module 18.

23  
24 The pose of each of the designated targets, derived  
25 from the "expected view" calculated by the modelling  
26 module 14, is fed into a transformation module 22  
27 together with the preliminary secondary images. The  
28 preliminary secondary images are transformed by the  
29 transformation module 22 so that they have the correct  
30 perspective appearance (size, shape and orientation) to  
31 match the corresponding target space as viewed by the  
32 camera 10.

33  
34 The original video image and the expected image  
35 calculated from the 3-d model 16 are both also passed  
36 to a matching module 24. The matching module 24

1 effectively superimposes the calculated expected image  
2 over the actual image as a basis for matching the two.  
3 It identifies as many as possible of the corners and  
4 edges of the target spaces corresponding to the  
5 designated targets and any other items of the venue  
6 model 16 present in the expected image. It uses these  
7 matches to refine the transformational match of the  
8 expected image to the actual image. Finally, the  
9 matcher extracts any foreground objects and lighting  
10 effects from the image areas of the designated targets.

11

12 The original primary image from the editing desk 12,  
13 the transformed secondary image and the output data  
14 from the matching module 24 are passed to one or more  
15 output modules 26 where they are combined to produce a  
16 final composite video output, in which the primary and  
17 secondary images are combined. There may be multiple  
18 output modules 26, each inserting different secondary  
19 images into the same primary images.

20

21 Obviously, for live transmission, this whole procedure  
22 has to happen in real time. Fortunately, the state of  
23 modern computing and image processing technology is  
24 such that the necessary hardware is not particularly  
25 expensive.

26

27 Each of the modules mentioned above is described in  
28 more detail below.

29

### 30 Camera Augmentation

31

32 Each camera is equipped with a device which  
33 continuously transmits additional camera data to the  
34 central station. This camera data could either be  
35 transmitted via a separate means such as additional  
36 cables or radio links, or could be incorporated into

1 the hidden parts of the video signal in the same way as  
2 teletext information. Methods and means for  
3 transmitting such data are well known.

4  
5 This camera data typically includes some or all of:  
6 a camera identifier;  
7 the camera position;  
8 the camera orientation;  
9 the lens focal length;  
10 the lens focusing distance;  
11 the camera aperture.

12  
13 The camera identifier is a string of characters which  
14 uniquely identifies each camera in use. The camera  
15 position is a set of three coordinate values giving the  
16 position of the camera in the coordinate system in use  
17 in the 3-d venue model. The camera orientation is  
18 another set of three values, defining the direction in  
19 which the camera is pointing. For example, this could  
20 be made up of three angles defining the camera viewing  
21 direction in the coordinate system used to define the  
22 camera position. The coordinate system used is not  
23 critical as long as all the cameras in use at a  
24 particular event supply the camera data in a way which  
25 is understood by the modelling and transformation  
26 modules.

27  
28 Since most cameras are fitted with zoom lenses, the  
29 lens focal length is required to define the scene for  
30 the purposes of secondary image transformation. The  
31 lens focusing distance and camera aperture are also  
32 required to define the scene for the purposes of  
33 transforming the secondary image in terms of which  
34 parts of the scene are in focus.

35  
36 The additional devices with which each camera is

1 equipped may depend on the role of the camera. For  
2 example, a particular camera may be fixed in position  
3 but adjustable in orientation. In this case, a  
4 calibration procedure may be used which results in an  
5 operator entering the camera's position into the device  
6 before the event starts. The orientation would be  
7 determined continuously by the device as would the  
8 focal length, focusing distance and aperture.

9

#### 10 The Venue Model

11

12 Key elements at the venue are represented within the  
13 general 3-d venue model 16.

14

15 The model may be based on a normal orthogonal 3-d  
16 coordinate system. The coordinate system origin used  
17 at a particular venue may be global or local in nature.  
18 For example, if the venue is a soccer stadium, it may  
19 be convenient to take the centre spot as the origin and  
20 to take the half-way line to define one axis direction,  
21 with an imaginary line running down the centre of the  
22 pitch defining a second axis direction. The third axis  
23 would then be a vertical line through the centre spot.

24

25 Each relevant permanent item of the venue is  
26 represented within the model in a way which  
27 encapsulates the item's important features for the  
28 purposes of the present system. Again, in the example  
29 of the soccer stadium, this could include:

30 the playing surface, represented as a planar  
31 surface with particular surface markings and a  
32 particular texture;

33 goalposts, represented as a solid object, for  
34 example, as the intersection of several cylindrical  
35 objects, having specific surface properties, e.g. white  
36 colour;

1 goal nets, which may be represented as an  
2 intersection of curvilinear objects with specific  
3 surface properties and having the property of  
4 flexibility;

5 advertising hoardings, which, in the simplest  
6 case, are represented as planar surfaces with complex  
7 surface properties, i.e. the physical advertisement  
8 (it is preferable that the surface properties are  
9 stored using a scale-invariant representation in order  
10 to simplify the matching process);

11 prominent permanent venue features: it is useful  
12 to the matching process if prominent features are  
13 included in the venue model; these may be stored as  
14 solid objects with surface properties (for example, if  
15 a grandstand contains a series of vertical pillars,  
16 then these could be used in the matching process to  
17 improve the accuracy of the process).

18

19 The methods and means for generating and using 3-d  
20 models, such as the venue model described above, and  
21 for determining the positions of objects within such  
22 models are all well known from other applications such  
23 as virtual reality modelling.

24

#### 25 Overall Signal Processing

26

27 The object of the signal processing performed by the  
28 system is to identify the position of the designated  
29 targets in the current image, to extract any foreground  
30 objects and lighting effects relevant to the designated  
31 targets, then to generate secondary images and insert  
32 them into the current primary image in place of the  
33 designated targets such that they look completely  
34 natural. The signal processing takes place in the  
35 following stages.

36

- 1     1.    Use the camera data in conjunction with the venue  
2     model to generate an expected image incorporating all  
3     the objects in the venue model which are expected to be  
4     seen in the actual image and to calculate the pose of  
5     each of the visible designated targets relative to the  
6     camera (modelling module 14).
- 7     2.    Identify as many as possible of the expected  
8     objects in the actual image (matching module 24).
- 9     3.    Use the individual item matches to refine the view  
10    details of the expected image (matching module).
- 11    4.    Project the borders of the designated targets onto  
12    the real image and refine the border positions, where  
13    appropriate with reference to edges and corners in the  
14    actual image (matching module 24).
- 15    5.    Match the expected designated target image to the  
16    corresponding region in the actual image, the match to  
17    be performed separately in colour space and intensity  
18    space. Any missing regions in the colour space match  
19    are assumed to be foreground objects. The bounding  
20    subregion of the target region is extracted and stored.  
21    The stored region includes colour and intensity  
22    information. Any mismatch regions occurring in  
23    intensity space only, e.g. shadows, which are not part  
24    of foreground objects are extracted and stored as  
25    intensity variations (matching module 24).
- 26    6.    Store the outcome of the matching process for use  
27    in matching the next frame.
- 28    7.    Transform the scale-invariant designated target  
29    model to fit the best estimate bounding region  
30    (transform module 22).
- 31    8.    Reassemble as many outgoing video signals as  
32    required by inserting the transformed secondary images  
33    into the original primary image and then reinserting  
34    foreground objects and lighting effects (output  
35    module).
- 36

1     Matching Module

2

3     The matching module 24 has several related functions.

4

5     The matcher first compares the expected view with the  
6     actual image to match corners and edges of items in the  
7     expected view with corresponding corners and edges in  
8     the actual image. This is greatly simplified by the  
9     fact that the expected image should be very close to  
10    the same view of the scene as the actual image. The  
11    object of this phase of matching is to correlate  
12    regions of the actual image with designated targets in  
13    the expected image. Corners are particularly  
14    beneficial in this part of the process since a corner  
15    match provides two constraints on the overall  
16    transformation whilst an edge match provides only one.  
17    Since the colour of the objects in the expected image  
18    is known from their representation in the venue model,  
19    this provides a further important clue in the matching  
20    process. When as many as possible of the corners and  
21    edges of the objects in the expected image have been  
22    matched to corners and edges in the actual image, a  
23    consistency check is carried out and any individual  
24    matches which are inconsistent with the overall  
25    transformation are rejected. Matching corners and  
26    edges in this way is a method well established in  
27    machine vision applications.

28

29    The outcome of the first phase of matching is a  
30    detailed mapping of the expected image onto the actual  
31    image. The second stage of matching is to deal with  
32    each designated target in turn to identify its exact  
33    boundary in the image and any foreground objects or  
34    lighting effects affecting the appearance of the  
35    corresponding physical object or area in the original  
36    image. This is done by using the corner and edge

1 matches and interpolating any missing sections of the  
2 boundary of the original object/area using the  
3 projected boundary of the designated target. For  
4 example, if the designated target is a rectangular  
5 advertising hoarding, then as long as sufficient  
6 segments of the boundary of the hoarding are  
7 identified, the position of the remaining segments can  
8 be calculated using the known segments and the known  
9 shape and size of the hoarding together with the known  
10 transformation into the image.

11  
12 The final stage of the matching process involves  
13 identifying foreground objects and lighting effects  
14 within the region of each designated target. This is  
15 based on transforming the scale invariant  
16 representation of the designated target in the venue  
17 model such that it fits exactly the bounding region of  
18 the corresponding ad in the original image. A match in  
19 colour space is then carried out within the bounding  
20 region to identify sections of the image which do not  
21 match the corresponding sections of the transformed  
22 model. These non-matching sections are taken to be  
23 foreground objects and these parts of the image are  
24 extracted and stored to be superimposed on top of the  
25 transformed secondary image in the final composite  
26 image. A match in intensity space is also carried out  
27 to identify intensity variations which are not part of  
28 the original object/area. These are considered to be  
29 lighting effects and an intensity transformation is  
30 used to extract these and keep them for later use in  
31 transforming the secondary image.

32  
33 Hence, the output from the matching process includes:  
34 the exact image boundary of all the designated  
35 targets;  
36 foreground objects in any of these regions; and



1 lighting effects in any of these regions.

2

3 Secondary Image Generation Module

4

5 One of the major advantages of using electronically  
6 generated secondary images rather than physical signs  
7 is in the extra scope for controlling the choice,  
8 positioning and content of the secondary image, e.g. an  
9 advertising message.

10

11 Generation of the secondary images uses a database 20  
12 of secondary image material. In addition to the actual  
13 secondary images, stored as scale-invariant  
14 representations, this database may include information  
15 such as:

16 the percentage of the available advertising space-  
17 time has been booked by each advertiser;

18 any preferences on which part of the event's  
19 duration and which part of the venue are to be used for  
20 each advertiser;

21 associations of particular secondary images with  
22 potential occurrences in the event being covered.

23

24 Another strength of the use of electronically  
25 integrated secondary images is the ability to generate  
26 different video outputs for different customers.

27 Hence, in an international event, different advertising  
28 material could be inserted into the video signal going  
29 to different countries. For example, say the USA is  
30 playing China at basketball. Most Americans don't read  
31 Chinese and most Chinese don't read English. So the  
32 transmission to China would include only advertisements  
33 in Chinese, while the broadcast in the USA would  
34 include only english language advertisements.

35

36 Generating a particular advertisement for display in

1 the present system may take place in the following  
2 stages:

3 choose the company whose advertisement will be  
4 displayed;

5 choose which of the selected company's  
6 advertisements is appropriate for the current context;

7 transform the stored representation of the  
8 selected advertisement to match the available region of  
9 the image.

10

11 For the first stage of this process, the selection of  
12 the advertiser, the destination of the video signal  
13 concerned is first determined. This indexes the  
14 advertisers for the output module 26 corresponding to  
15 that destination. Next, a check is made to see how  
16 much advertising time each advertiser has had during  
17 the event so far relative to how much they have booked.  
18 The advertiser is selected on this basis, taking  
19 account of advertiser preferences such as location and  
20 timing.

21

22 The next stage, the selection of one advertisement from  
23 a set supplied by the advertiser to replace a  
24 designated target in the original image, is based on  
25 factors including:

26 the size of the space available;

27 the location of the designated target;

28 the phase of the event;

29 any notable occurrences during the event.

30

31 For example, an advertiser may choose to supply some  
32 advertisements containing a lot of detail and some  
33 which are very simple. If the space available is  
34 large, perhaps because the camera concerned is showing  
35 a close up of a soccer player about to take a corner  
36 and the advertising space available fills a large part

1 of the image, then it may be appropriate to fit a more  
2 detailed advertisement where the details will be  
3 visible. At the other extreme, if a particular camera  
4 is showing a long view, then it may be better to select  
5 a very simple advertisement with strong graphics so  
6 that the advertisement is legible on the screen.

7  
8 Note also that the selection of advertisements can be  
9 influenced by what has happened in the event. For  
10 example, say a particular player, X, has just scored a  
11 goal. Then an advertiser who manufactures drink, Y, may  
12 want to display something to the effect that "X drinks  
13 Y". To meet this need the system has the capability to  
14 store advertisements which are only active (i.e.  
15 available for selection) when a particular event has  
16 taken place. Additionally, these advertisements can  
17 have place holders where the name of a participant or  
18 some other details can be entered when the ad is made  
19 active. This could be useful if drinks advertiser Y  
20 has a contract with a whole team. Then when any team  
21 member does something exceptional, that team member's  
22 name, or other designation, could be inserted into the  
23 advertisement.

24  
25 Note also that there is no restriction on  
26 advertisements being static. As long as the  
27 advertisement still looked as though it was part of the  
28 event, it could be completely dynamic. For example, an  
29 advertising video could be inserted into a suitable  
30 designated target. One particular case might be where  
31 the venue concerned has a large playback screen, such  
32 as at many cricket and athletics events. The screen  
33 would be used to show replays of the event to the  
34 spectators present, but it could also be a designated  
35 target for the present system. Such a screen would  
36 then be a good candidate for showing video advertising

1 material.

2

3 A further aspect of the process of secondary image  
4 generation relates to how to change images. Clearly,  
5 if a camera is panning, then different secondary images  
6 can be included as different parts of the venue come  
7 into the image. Note that it is important to record  
8 which secondary image is being displayed on which  
9 designated target, since a cut from one camera to  
10 another should not cause the secondary image to change  
11 if the two cameras are capturing the same designated  
12 target. It can also occur that one camera will be used  
13 for a particularly long time and it and it may be  
14 desirable to change the secondary images in the  
15 composite image part way through the shot. This is  
16 accomplished by simulating the change of a physical ad.  
17 For example, there are physical advertising hoardings  
18 available which are able to show more than one ad,  
19 either by rotating a strip containing the ads or by  
20 rotating some triangular segments, each of whose faces  
21 contains portions of different ads. To change a  
22 secondary image while it is in shot, the secondary  
23 image generation process may simulate the operation of  
24 a physical hoarding, for example, by appearing to  
25 rotate segments of a hoarding to switch from one ad to  
26 the next.

27

#### 28 Transform Module

29

30 The pose of the physical advertising space relative to  
31 the camera concerned is known from the additional  
32 camera data and the 3-d venue model 16. Hence,  
33 transforming the scale-invariant representation of the  
34 chosen secondary image into a 2-d image region with the  
35 correct perspective appearance is a straightforward  
36 task. In addition to the pose being correct, the

1 secondary image has to fit the target space exactly.  
2 The region bounding the space is supplied by the  
3 matching process. Hence, transforming the ad involves:  
4 using the additional camera data and 3-d venue  
5 model 16 to calculate the perspective appearance of the  
6 secondary image (this is done in the modelling module  
7 14);.

8 using the matching information to scale the  
9 secondary image to fit the space available.

10  
11 The secondary image is now ready to be dropped into the  
12 original video image.

13  
14 Output Module

15  
16 One output module 26 is required for each outgoing  
17 video signal. Hence, if the final of the World Cup is  
18 being transmitted to 100 countries which have been  
19 split into 10 areas for advertising, then ten output  
20 modules would be required.

21  
22 The output module 26 takes one set of secondary images  
23 and inserts them into the original primary image. It  
24 then takes the foreground object and lighting effects  
25 generated by the matching process and reintegrates  
26 them. In the case of the foreground objects, this  
27 requires parts of the inserted secondary images to be  
28 overwritten with the foreground objects. In the case  
29 of lighting effects, such as shadows, the image  
30 segments containing the secondary image must be  
31 modified such that the secondary image looks as if it  
32 is subjected to the same lighting effects as the  
33 corresponding part of the original scene. This is done  
34 by separating out the colour and intensity information  
35 and modifying them appropriately. Methods for doing  
36 this are well known in the field of computer graphics.

1 Use of the present invention has many benefits for  
2 advertisers, particularly at large international  
3 events. Some of these benefits are as follows:

4 different advertisements can be shown in different  
5 countries or regions thereby improving targeting and  
6 making sure that the advertising regulations of  
7 individual countries, e.g. with respect to alcohol and  
8 tobacco, are not violated;

9 each advertiser can be guaranteed a percentage of  
10 the total exposure;

11 the detail of the advertisements can be adjusted  
12 automatically based on their size in the TV image to  
13 improve their legibility and impact;

14 there may be much greater creative scope in the  
15 design of the advertisements;

16 by recording some extra information with the  
17 individual camera video signals, different  
18 advertisements can be used in subsequent use of the  
19 original footage: for example, different advertisements  
20 could be used in programmes of highlights than in live  
21 broadcasts, and different advertisements again could be  
22 used in subsequent video products.

23

24 Systems for replacing parts of video images with parts  
25 of other images such that the replacement parts appear  
26 to be a natural part of the original image are known in  
27 the prior art. However, the systems described in the  
28 prior art have serious limitations which are overcome  
29 by the present invention.

30

31 One area of the prior art is based on colour or chroma  
32 keying. This depends on being able to control the  
33 colour of everything in the image and is not practical  
34 as a general purpose system.

35

36 Another area of prior art involves a human operator

1 manually selecting the areas to be replaced and  
2 performing various functions to deal with foreground  
3 objects and lighting effects. This method is very time  
4 consuming and expensive and obviously not applicable to  
5 live broadcasts.

6  
7 Another area of prior art specifies automatic  
8 replacement of an advertising logo using the pose of  
9 the identified logo to transform the virtual ad  
10 (WO93/06691). However, this method does not describe  
11 any way of dealing with foreground objects or lighting  
12 effects.

13  
14 The main advantages of the present invention over the  
15 prior art are considered to be:

16 augmentation of cameras and the use of a full 3-d  
17 venue model to enable generation of an expected image  
18 and reliable and fast matching of the expected image to  
19 an actual image without relying on colour keying or  
20 extensive searching or analysis of the actual image;

21 use of the full 3-d venue model together with the  
22 additional camera data to eliminate the need to  
23 estimate the pose of physical ads from the image data;

24 separation of the video signal into colour and  
25 intensity images for separate treatment of foreground  
26 objects and lighting effects;

27 use of corner and edge detection and matching as  
28 the basis for superimposing expected image segments  
29 over actual image segments;

30 use of stored scale-invariant representations of  
31 the physical designated targets to greatly simplify  
32 identification of foreground objects and lighting  
33 effects.

34  
35 As a result of these improvements, the present  
36 invention is much more generally applicable than those

1     based on the prior art.  
2  
3     Improvements and modifications may be incorporated  
4     without departing from the scope of the invention.  
5



1     Claims

2

3     1.    A method of modifying a first video image of a  
4     real world scene to include a second video image, such  
5     that said second image appears to be superimposed on  
6     the surface of an object appearing within said first  
7     image, wherein said second image is derived by  
8     transforming a preliminary second image to match the  
9     size, shape and orientation of said surface as seen in  
10    said first image and said second image is combined with  
11    said first image to produce a composite final image;  
12    said method including:

13        a preliminary step of constructing a three-  
14    dimensional computer model of the environment  
15    containing the real world scene, said model including  
16    at least one target space within said environment upon  
17    which said second image is to be superimposed;

18        generating camera data defining at least the  
19    location, orientation and focal length of a camera  
20    generating said first image; and

21        transforming the preliminary second image on the  
22    basis of said model and said camera data so as to match  
23    said target space as seen in the first image, prior to  
24    combining said first image and said second image.

25

26    2.    A method as claimed in Claim 1, wherein the  
27    transformation of the preliminary second image includes  
28    manipulation thereof to take account of lighting  
29    conditions in the image of the real world scene.

30

31    3.    A method as claimed in Claim 2, wherein objects  
32    included in said model are matched with corresponding  
33    regions of said first image, intensity information  
34    relating to matched objects is compared with intensity  
35    information relating to said corresponding image  
36    region, regions of intensity mismatch within said

1 corresponding regions are identified as lighting  
2 variations and, when said second image is transformed,  
3 the intensity of portions thereof is varied on the  
4 basis of said regions of intensity mismatch so as to  
5 simulate lighting variations within the first image.  
6

7 4. A method as claimed in any preceding Claim,  
8 wherein the combination of the first and second images  
9 includes manipulation thereof to take account of  
10 foreground objects in the image of the real world  
11 scene.  
12

13 5. A method as claimed in Claim 4, wherein objects  
14 included in said model are matched with corresponding  
15 regions of said first image, colour information  
16 relating to matched objects is compared with colour  
17 information relating to said corresponding image  
18 region, regions of colour mismatch within said  
19 corresponding regions are identified as foreground  
20 objects and, when said first and second images are  
21 combined, said first image is retained in preference to  
22 said second image within said colour mismatch regions.  
23

24 6. A method as claimed in any preceding Claim,  
25 wherein said camera data and said computer model are  
26 combined to compute a representation of the image  
27 expected from the camera.  
28

29 7. A method as claimed in Claim 6, wherein features  
30 of said expected image are matched with features of  
31 said first image.  
32

33 8. A method as claimed in Claim 7, wherein said  
34 matching of the expected image and the first image is  
35 used to refine the boundary of the target space within  
36 the expected image.

- 1     9. A method as claimed in Claim 8, wherein the  
2     transformation of the shape, size and orientation of  
3     the preliminary second image is based on said refined  
4     target boundary.  
5
- 6     10. A method as claimed in Claim 7, Claim 8 or Claim  
7     9, wherein said matching of the expected image and the  
8     first image includes comparison of colour and intensity  
9     information for the purpose of identifying foreground  
10    objects and lighting variations in said first image.  
11
- 12    11. A method as claimed in any one of Claims 7 to 10,  
13    wherein said first image and said second image are  
14    combined on the basis of said matching of features  
15    between the expected image and the first image.  
16
- 17    12. A method as claimed in any one of Claims 7 to 11,  
18    wherein said computer model includes scale-invariant  
19    colour representations of surface properties of said  
20    target spaces and said expected image incorporates said  
21    colour representations of said target spaces.  
22
- 23    13. A method as claimed in any preceding Claim,  
24    wherein said first video image is a live action video  
25    image and said composite image is generated in real  
26    time.  
27
- 28    14. A method as claimed in any preceding Claim wherein  
29    multiple second images are superimposed upon multiple  
30    target spaces.  
31
- 32    15. A method as claimed in any preceding Claim,  
33    wherein multiple composite images are generated, each  
34    comprising the same first image combined with differing  
35    second images.  
36

1 16. A method as claimed in any preceding Claim,  
2 wherein said second image is selected automatically  
3 from a plurality of images, in accordance with  
4 predetermined selection criteria.

5  
6 17. A method as claimed in any preceding Claim,  
7 wherein said first image is selected from a plurality  
8 of video images generated by a plurality of cameras.

9  
10 18. Apparatus for generating a composite video image  
11 comprising a combination of a first video image of a  
12 real world scene and a second video image, such that  
13 said second image appears to be superimposed on the  
14 surface of an object appearing within said first image,  
15 including:

16 at least one camera for generating said first  
17 image;

18 means for generating said second image by  
19 transforming a preliminary second image to match the  
20 size, shape and orientation of said surface as seen in  
21 said first image; and

22 means for combining said second image with said  
23 first image to produce a composite final image;

24 said apparatus including:

25 means for storing a three-dimensional computer  
26 model of the environment containing the real world  
27 scene, said model including at least one target space  
28 within said environment upon which said second image is  
29 to be superimposed;

30 means for generating camera data defining at least  
31 the location, orientation and focal length of a camera  
32 generating said first image; and

33 means for transforming the preliminary second  
34 image on the basis of said model and said camera data  
35 so as to match said target space as seen in the first  
36 image, prior to combining said first image and said

1 second image.

2

3 19. Apparatus as claimed in Claim 18, wherein the  
4 means for transforming the preliminary second image  
5 includes means for manipulating said second image to  
6 take account of lighting conditions in the first image  
7 of the real world scene.

8

9 20. Apparatus as claimed in Claim 19, means for  
10 matching objects included in said model with  
11 corresponding regions of said first image, said  
12 matching means including means for comparing intensity  
13 information relating to matched objects with intensity  
14 information relating to said corresponding image  
15 region, and means for identifying regions of intensity  
16 mismatch within said corresponding regions, and wherein  
17 said image transforming means includes means for  
18 varying the intensity of portions of said second image  
19 on the basis of said regions of intensity mismatch so  
20 as to simulate lighting variations within the first  
21 image.

22

23 21. Apparatus as claimed in any one of Claims 18 to  
24 20, wherein the means for combining the first and  
25 second images includes means for manipulating said  
26 second image to take account of foreground objects in  
27 the image of the real world scene.

28

29 22. Apparatus as claimed in Claim 21, including means  
30 for matching objects included in said model with  
31 corresponding regions of said first image, said  
32 matching means including means for comparing colour  
33 information relating to matched objects with colour  
34 information relating to said corresponding image  
35 region, and means for identifying regions of colour  
36 mismatch within said corresponding regions, and wherein

1    said image combining means includes means for  
2    manipulating said second image such that, when said  
3    first and second images are combined, said first image  
4    is retained in preference to said second image within  
5    said colour mismatch regions.

6  
7    23. Apparatus as claimed in any one of Claims 18 to  
8    22, including computer modelling means adapted to  
9    compute a representation of the image expected from the  
10   camera on the basis of said camera data and said  
11   computer model.

12  
13   24. Apparatus as claimed in Claim 23, including means  
14   for matching features of said expected image with  
15   features of said first image.

16  
17   25. Apparatus as claimed in Claim 24, wherein said  
18   means for matching the expected image and the first  
19   image is further adapted to refine the boundary of the  
20   target space within the expected image.

21  
22   26. Apparatus as claimed in Claim 25, wherein the  
23   image transformation means is adapted to effect  
24   transformation of the shape, size and orientation of  
25   the preliminary second image based on said refined  
26   target boundary.

27  
28   27. Apparatus as claimed in Claim 24, Claim 25 or  
29   Claim 26, wherein said means for matching the expected  
30   image and the first image includes means for comparing  
31   colour and intensity information for the purpose of  
32   identifying foreground objects and lighting variations  
33   in said first image.

34  
35   28. Apparatus as claimed in any one of Claims 24 to  
36   27, wherein said means for combining said first image

1 and said second image are adapted to effect said  
2 combination on the basis of said matching of features  
3 between the expected image and the first image.  
4

5 29. Apparatus as claimed in any one of Claims 24 to  
6 28, wherein said computer model includes scale-  
7 invariant colour representations of surface properties  
8 of said target spaces and said modelling means is  
9 adapted to generate expected images incorporating said  
10 colour representations of said target spaces.  
11

12 30. Apparatus as claimed in any one of Claims 18 to  
13 29, wherein said first video image is a live action  
14 video image and the apparatus is adapted to generate  
15 said composite image in real time.  
16

17 31. Apparatus as claimed in any one of Claims 18 to  
18 30, wherein the apparatus is adapted to superimpose  
19 multiple second images upon multiple target spaces.  
20

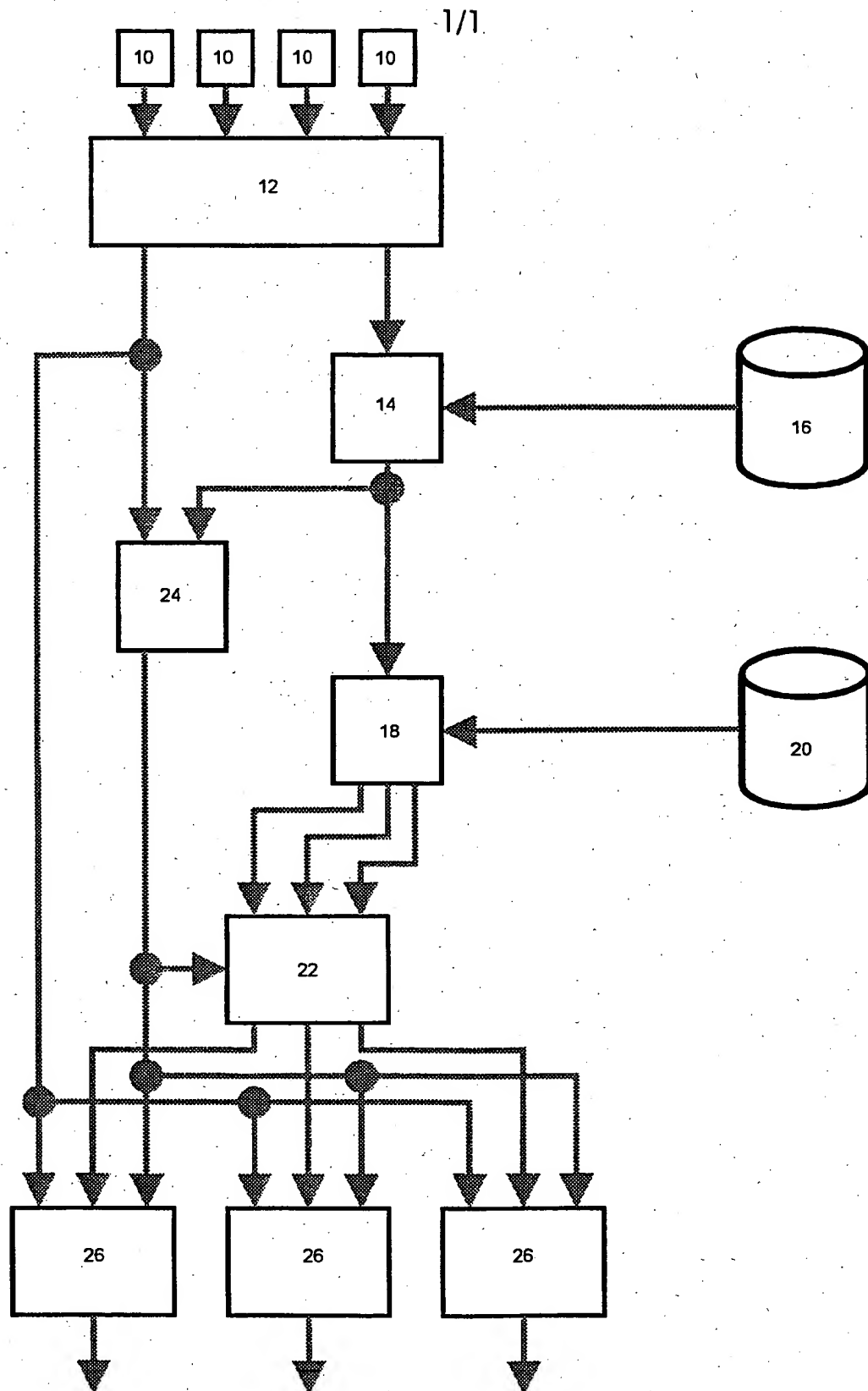
21 32. Apparatus as claimed in any one of Claims 18 to  
22 31, including multiple output means, each of said  
23 output means being adapted to generate different  
24 composite images, each of said different composite  
25 images comprising the same first image combined with  
26 differing second images.  
27

28 33. Apparatus as claimed in any one of Claims 18 to  
29 32, including means for storing a plurality of images  
30 and means for automatically selecting said second image  
31 from said plurality of images, in accordance with  
32 predetermined selection criteria.  
33

34 34. Apparatus as claimed in any one of Claims 18 to  
35 33, wherein a plurality of cameras are connected to  
36 video editing means and said first image is selected

- 1 from a plurality of video images generated by said
- 2 plurality of cameras.
- 3





SUBSTITUTE SHEET (RULE 26)

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/GB 96/01682

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 H04N5/272 H04N5/262

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,95 10919 (ORAD INC ;SHARIR AVI (IL); TAMIR MICHAEL (IL)) 20 April 1995 see page 3, line 4 - page 11, line 16 see page 15, line 4 - page 36, line 30; figures 1-18	1-34
X	WO,A,93 06691 (SARNOFF DAVID RES CENTER) 1 April 1993 cited in the application	1,2,4,6, 7,10,11, 13,14, 17-19, 21,23, 24,27, 28,30, 31,34 3,5,9, 20,22,26
A	see page 7, line 34 - page 21, line 2; figures 4-7	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

7 November 1996

Date of mailing of the international search report

4.12.96

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

De Paepe, W

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/GB 96/01682

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	US,A,5 491 517 (KREITMAN HAIM ET AL) 13 February 1996	1,2,4,6, 7,10,11, 13, 17-19, 21,23, 24,27, 28,30,34
P,A	see column 1, line 59 - column 3, line 44 see column 5, line 51 - column 6, line 2; figure 2 see column 6, line 50 - column 6, line 64; figures 1,2 see column 7, line 42 - column 8, line 54; figures 5-9 see column 12, line 59 - column 14, line 26; figures 18-21 -----	3,5,9, 14,20, 22,26, 31-33

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 96/01682

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9510919	20-04-95	AU-A- 6298794 BR-A- 9406756 CN-A- 1119481 EP-A- 0683961	04-05-95 02-04-96 27-03-96 29-11-95
WO-A-9306691	01-04-93	AU-B- 663731 AU-A- 2672092 CA-A- 2119272 EP-A- 0604572 JP-T- 6510893 US-A- 5566251	19-10-95 27-04-93 01-04-93 06-07-94 01-12-94 15-10-96
US-A-5491517	13-02-96	AU-A- 1933495 CA-A- 2179031 WO-A- 9525399	03-10-95 21-09-95 21-09-95